PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/53, 15/82, 5/10, C12P 7/64, C11B 1/00, A61K 31/20, A23L 1/30, A23K 1/00

A1

(11) International Publication Number: WO 98/46764

(43) International Publication Date: 22 October 1998 (22.10.98)

(21) International Application Number: PCT/US98/07421

(22) International Filing Date: 10 April 1998 (10.04.98)

(30) Priority Data:

08/833,610	11 April 1997 (11.04.97)	US
08/834,033	11 April 1997 (11.04.97)	US
08/834,655	11 April 1997 (11.04.97)	US
08/956,985	24 October 1997 (24.10.97)	US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications

US	08/834,655 (CIP)
Filed on	11 April 1997 (11.04.97)
US	08/833,610 (CIP)
Filed on	11 April 1997 (11.04.97)
US	08/834,033 (CIP)
Filed on	11 April 1997 (11.04.97)
US	08/956,985 (CIP)
Filed on	24 October 1997 (24.10.97)

(71) Applicants *(for all designated States except US)*: CALGENE LLC [US/US]; 1920 Fifth Street, Davis, CA 95616 (US). ABBOTT LABORATORIES [US/US]; 100 Abbott Park Road, Abbott Park, IL 60064–3500 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: KNUTZON, Deborah [US/US]; 6110 Rockhurst Way, Granite Bay, CA 95746 (US). MUKERJI, Pradip [US/US]; 1069 Arcaro Drive, Gahanna, OH 43230 (US). HUANG, Yung–Sheng [–/US]; 2462 Danvers Court, Upper Arlington, OH 43220 (US). THURMOND, Jennifer [US/US]; 3702 Adirondack, Columbus, OH 43231 (US). CHAUDHARY, Sunita [IN/US]; 3419 Woodbine Place, Pearland, TX 77584 (US). LEONARD, Amanda, Eun–Yeong [US/US]; 581 Shadewood Court, Gahanna, OH 43230 (US).

(74) Agents: WARD, Michael, R. et al.; Limbach & Limbach L.L.P., 2001 Ferry Building, San Francisco, CA 94111–4262 (US).

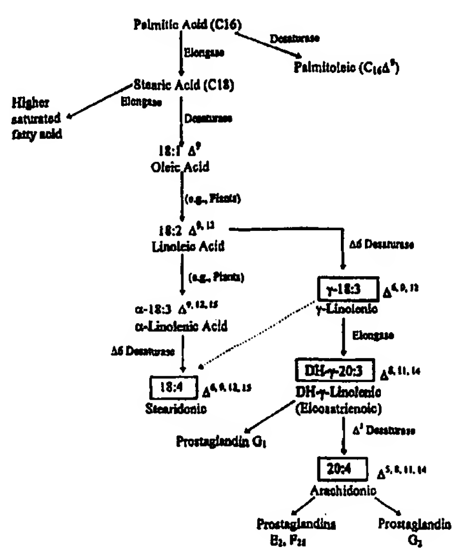
(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: METHODS AND COMPOSITIONS FOR SYNTHESIS OF LONG CHAIN POLYUNSATURATED FATTY ACIDS IN PLANTS

(57) Abstract

The present invention relates to compositions and methods for preparing polyunsaturated long chain fatty acids in plants, plant parts and plant cells, such as leaves, roots, fruits and seeds. Nucleic acid sequences and constructs encoding fatty acid desaturases, including Δ5–desaturases, Δ6–desaturases and Δ12–desaturases, are used to generate transgenic plants, plant parts and cells which contain and express one or more transgenes encoding one or more desaturases. Expression of the desaturases with different substrate specificities in the plant system permit the large scale production of polyunsaturated long chain fatty acids such as docosahexaenoic acid, eicosapentaenoic acid, α–linolenic acid, gamma–linolenic acid, arachidonic acid and the like for modification of the fatty acid profile of plants, plant parts and tissues. Manipulation of the fatty acid profiles allows for the production of commercial quantities of novel plant oils and products.



comprising a transcriptional and translational initiation regulatory region, joined in reading frame 5' to a DNA sequence encoding a desaturase polypeptide capable of modulating the production of PUFAs. Expression of the desaturase polypeptide provides for an alteration in the PUFA profile of host plant cells as a result of altered concentrations of enzymes involved in PUFA biosynthesis. Of particular interest is the selective control of PUFA production in plant tissues and/or plant parts such as leaves, roots, fruits and seeds. The invention finds use for example in the large scale production of DHA, EPA, ARA, and GLA and for modification of the fatty acid profile of edible plant tissues and/or plant parts.

The present invention further includes a purified nucleotide sequence or polypeptide sequence that is substantially related or homologous to the nucleotide and peptide sequences presented in SEQ ID NO:1 - SEQ ID NO:52. The present invention is further directed to methods of using the sequences presented in SEQ ID NO:1 to SEQ ID NO:40 as probes to identify related sequences, as components of expression systems and as components of systems useful for producing transgenic oil.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows possible pathways for the synthesis of arachidonic acid (20:4 Δ5, 8, 11, 14) and stearidonic acid (18:4 Δ6, 9, 12, 15) from palmitic acid (C_{16}) from a variety of organisms, including algae, *Mortierella* and humans. These PUFAs can serve as precursors to other molecules important for humans and other animals, including prostacyclins, leukotrienes, and prostaglandins, some of which are shown.

Figure 2 shows possible pathways for production of PUFAs in addition to ARA, including EPA and DHA, again compiled from a variety of organisms.

Figure 3A-E shows the DNA sequence (SEQ ID NO:1) of the *Mortierella alpina* Δ6 desaturase and the deduced amino acid sequence (SEQ ID NO:2).

Figure 4 shows an alignment of the *Mortierella alpina* Δ6 desaturase amino acid sequence with other Δ6 desaturases and related sequences (SEQ ID NOS:7, 8, 9, 10, 11, 12 and 13).

Figure 5A-D shows the DNA sequence of the *Mortierella alpina* Δ12 desaturase (SEQ ID NO:3) and the deduced amino acid sequence (SEQ ID NO:4)

Figure 6 shows the deduced amino acid sequence (SEQ ID NO:14) of the PCR fragment (see Example 1).

Figure 7A-D shows the DNA sequence of the *Mortierella alpina* Δ5 desaturase (SEQ ID NO:5).

Figure 8 shows alignments of the protein sequence of the Δ5 desaturase (SEQ ID NO:6) with Δ6 desaturases and related sequences (SEQ ID NOS:15, 16, 17, 18).

Figure 9 shows alignments of the protein sequence of the Ma 29 and contig 253538a.

Figure 10 shows alignments of the protein sequence of Ma 524 and contig 253538a.

BRIEF DESCRIPTION OF THE SEQUENCE LISTINGS

SEQ ID NO:1 shows the DNA sequence of the *Mortierella alpina* Δ6 desaturase.

SEQ ID NO:2 shows the amino acid sequence of the *Mortierella alpina* Δ6 desaturase.

SEQ ID NO:3 shows the DNA sequence of the *Mortierella alpina* Δ12 desaturase.

SEQ ID NO:4 shows the amino acid sequence of the *Mortierella alpina* Δ12 desaturase.

SEQ ID NO:5 shows the DNA sequence of the *Mortierella alpina* Δ5 desaturase.

SEQ ID NO:6 shows the amino acid sequence *Mortierella alpina* Δ5 desaturase.

SEQ ID NO:7 - SEQ ID NO:13 show amino acid sequences that relate to *Mortierella alpina* Δ6 desaturase.

SEQ ID NO:14 shows an amino acid sequence of a PCR fragment of Example 1.

SEQ ID NO:15 - SEQ ID NO:18 show amino acid sequences that relate to *Mortierella alpina* Δ5 and Δ6 desaturases.

SEQ ID NO:19 - SEQ ID NO:30 show PCR primer sequences.

SEQ ID NO:31 - SEQ ID NO:37 show human nucleotide sequences.

SEQ ID NO:38 - SEQ ID NO:44 show human peptide sequences.

SEQ ID NO:45 - SEQ ID NO:46 show the nucleotide and amino acid sequence of a *Dictyostelium discoideium* desaturase.

SEQ ID NO:47 - SEQ ID NO:50 show the nucleotide and deduced amino acid sequence of a *Schizochytrium* cDNA clone.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

In order to ensure a complete understanding of the invention, the following definitions are provided:

Δ5-Desaturase: Δ5 desaturase is an enzyme which introduces a double bond between carbons 5 and 6 from the carboxyl end of a fatty acid molecule.

Δ6-Desaturase: Δ6-desaturase is an enzyme which introduces a double bond between carbons 6 and 7 from the carboxyl end of a fatty acid molecule.

Δ9-Desaturase: Δ9-desaturase is an enzyme which introduces a double bond between carbons 9 and 10 from the carboxyl end of a fatty acid molecule.

```
                                                                     60
                                                                      *
CGACACTCCT TCCTTCTTCT CACCCGTCCT AGTCCCCTTC AACCCCCCTC TTTGACAAAG

ACAACAAACC ATG GCT GCT GCT CCC AGT GTG AGG ACG TTT ACT CGG GCC GAG
           Met Ala Ala Ala Pro Ser Val Arg Thr Phe Thr Arg Ala Glu

       120
        *
GTT TTG AAT GCC GAG GCT CTG AAT GAG GGC AAG AAG GAT GCC GAG GCA
Val Leu Asn Ala Glu Ala Leu Asn Glu Gly Lys Lys Asp Ala Glu Ala

                       180
                        *
CCC TTC TTG ATG ATC ATC GAC AAC AAG GTG TAC GAT GTC CGC GAG TTC
Pro Phe Leu Met Ile Ile Asp Asn Lys Val Tyr Asp Val Arg Glu Phe

                                        240
                                         *
GTC CCT GAT CAT CCC GGT GGA AGT GTG ATT CTC ACG CAC GTT GGC AAG
Val Pro Asp His Pro Gly Gly Ser Val Ile Leu Thr His Val Gly Lys

                                                        300
                                                          *
GAC GGC ACT GAC GTC TTT GAC ACT TTT CAC CCC GAG GCT GCT TGG GAG
Asp Gly Thr Asp Val Phe Asp Thr Phe His Pro Glu Ala Ala Trp Glu

ACT CTT GCC AAC TTT TAC GTT GGT GAT ATT GAC GAG AGC GAC CGC GAT
Thr Leu Ala Asn Phe Tyr Val Gly Asp Ile Asp Glu Ser Asp Arg Asp

       360
        *
ATC AAG AAT GAT GAC TTT GCG GCC GAG GTC CGC AAG CTG CGT ACC TTG
Ile Lys Asn Asp Asp Phe Ala Ala Glu Val Arg Lys Leu Arg Thr Leu
```

FIG. 3A

```
                                420
                                 *
TTC CAG TCT CTT GGT TAC TAC GAT TCT TCC AAG GCA TAC TAC GCC TTC
Phe Gln Ser Leu Gly Tyr Tyr Asp Ser Ser Lys Ala Tyr Tyr Ala Phe

                                        480
                                         *
AAG GTC TCG TTC AAC CTC TGC ATC TGG GGT TTG TCG ACG GTC ATT GTG
Lys Val Ser Phe Asn Leu Cys Ile Trp Gly Leu Ser Thr Val Ile Val

                                                        540
                                                         *
GCC AAG TGG GGC CAG ACC TCG ACC CTC GCC AAC GTG CTC TCG GCT GCG
Ala Lys Trp Gly Gln Thr Ser Thr Leu Ala Asn Val Leu Ser Ala Ala

CTT TTG GGT CTG TTC TGG CAG CAG TGC GGA TGG TTG GCT CAC GAC TTT
Leu Leu Gly Leu Phe Trp Gln Gln Cys Gly Trp Leu Ala His Asp Phe

        600
         *
TTG CAT CAC CAG GTC TTC CAG GAC CGT TTC TGG GGT GAT CTT TTC GGC
Leu His His Gln Val Phe Gln Asp Arg Phe Trp Gly Asp Leu Phe Gly

                                660
                                 *
GCC TTC TTG GGA GGT GTC TGC CAG GGC TTC TCG TCC TCG TGG TGG AAG
Ala Phe Leu Gly Gly Val Cys Gln Gly Phe Ser Ser Ser Trp Trp Lys

                                        720
                                         *
GAC AAG CAC AAC ACT CAC CAC GCC GCC CCC AAC GTC CAC GGC GAG GAT
Asp Lys His Asn Thr His His Ala Ala Pro Asn Val His Gly Glu Asp

                                                        780
                                                         *
```

FIG. 3B

```
CCC GAC ATT GAC ACC CAC CCT CTG TTG ACC TGG AGT GAG CAT GCG TTG
Pro Asp Ile Asp Thr His Pro Leu Leu Thr Trp Ser Glu His Ala Leu

GAG ATG TTC TCG GAT GTC CCA GAT GAG GAG CTG ACC CGC ATG TGG TCG
Glu Met Phe Ser Asp Val Pro Asp Glu Glu Leu Thr Arg Met Trp Ser

        840
         *
CGT TTC ATG GTC CTG AAC CAG ACC TGG TTT TAC TTC CCC ATT CTC TCG
Arg Phe Met Val Leu Asn Gln Thr Trp Phe Tyr Phe Pro Ile Leu Ser

                            900
                             *
TTT GCC CGT CTC TCC TGG TGC CTC CAG TCC ATT CTC TTT GTG CTG CCT
Phe Ala Arg Leu Ser Trp Cys Leu Gln Ser Ile Leu Phe Val Leu Pro

                                            960
                                             *
AAC GGT CAG GCC CAC AAG CCC TCG GGC GCG CGT GTG CCC ATC TCG TTG
Asn Gly Gln Ala His Lys Pro Ser Gly Ala Arg Val Pro Ile Ser Leu

                                                        1020
                                                           *
GTC GAG CAG CTG TCG CTT GCG ATG CAC TGG ACC TGG TAC CTC GCC ACC
Val Glu Gln Leu Ser Leu Ala Met His Trp Thr Trp Tyr Leu Ala Thr

ATG TTC CTG TTC ATC AAG GAT CCC GTC AAC ATG CTG GTG TAC TTT TTG
Met Phe Leu Phe Ile Lys Asp Pro Val Asn Met Leu Val Tyr Phe Leu

       1080
         *
GTG TCG CAG GCG GTG TGC GGA AAC TTG TTG GCG ATC GTG TTC TCG CTC
Val Ser Gln Ala Val Cys Gly Asn Leu Leu Ala Ile Val Phe Ser Leu
```

FIG. 3C

```
                        1140
                          *
AAC CAC AAC GGT ATG CCT GTG ATC TCG AAG GAG GAG GCG GTC GAT ATG
Asn His Asn Gly Met Pro Val Ile Ser Lys Glu Glu Ala Val Asp Met

                                        1200
                                          *
GAT TTC TTC ACG AAG CAG ATC ATC ACG GGT CGT GAT GTC CAC CCG GGT
Asp Phe Phe Thr Lys Gln Ile Ile Thr Gly Arg Asp Val His Pro Gly

                                                        1260
                                                          *
CTA TTT GCC AAC TGG TTC ACG GGT GGA TTG AAC TAT CAG ATC GAG CAC
Leu Phe Ala Asn Trp Phe Thr Gly Gly Leu Asn Tyr Gln Ile Glu His

CAC TTG TTC CCT TCG ATG CCT CGC CAC AAC TTT TCA AAG ATC CAG CCT
His Leu Phe Pro Ser Met Pro Arg His Asn Phe Ser Lys Ile Gln Pro

    1320
      *
GCT GTC GAG ACC CTG TGC AAA AAG TAC AAT GTC CGA TAC CAC ACC ACC
Ala Val Glu Thr Leu Cys Lys Lys Tyr Asn Val Arg Tyr His Thr Thr

                        1380
                          *
GGT ATG ATC GAG GGA ACT GCA GAG GTC TTT AGC CGT CTG AAC GAG GTC
Gly Met Ile Glu Gly Thr Ala Glu Val Phe Ser Arg Leu Asn Glu Val

                                        1440
                                          *
TCC AAG GCT GCC TCC AAG ATG GGT AAG GCG CAG TAAAAAAAA AAACAAGGAC
Ser Lys Ala Ala Ser Lys Met Gly Lys Ala Gln
```

FIG. 3D

```
                                             1500
                                                *
GTTTTTTTTC GCCAGTGCCT GTGCCTGTGC CTGCTTCCCT TGTCAAGTCG AGCGTTTCTG
                                             1560
                                                *
GAAAGGATCG TTCAGTGCAG TATCATCATT CTCCTTTTAC CCCCCGCTCA TATCTCATTC
ATTTCTCTTA TTAAACAACT TGTTCCCCCC TTCACCG
```

FIG. 3E

```
                                                                  60
                                                                   *
GTCCCCTGTC GCTGTCGGCA CACCCCATCC TCCCTCGCTC CCTCTGCGTT TGTCCTTGGC
                                                                 120
                                                                   *
CCACCGTCTC TCCTCCACCC TCCGAGACGA CTGCAACTGT AATCAGGAAC CGACAAATAC
                                                                  180
                                                                   *
ACGATTTCTT TTTACTCAGC ACCAACTCAA AATCCTCAAC CGCAACCCTT TTTCAGG ATG
                                                               Met

GCA CCT CCC AAC ACT ATC GAT GCC GGT TTG ACC CAG CGT CAT ATC AGC
Ala Pro Pro Asn Thr Ile Asp Ala Gly Leu Thr Gln Arg His Ile Ser
            240
             *
ACC TCG GCC CCA AAC TCG GCC AAG CCT GCC TTC GAG CGC AAC TAC CAG
Thr Ser Ala Pro Asn Ser Ala Lys Pro Ala Phe Glu Arg Asn Tyr Gln
                                300
                                 *
CTC CCC GAG TTC ACC ATC AAG GAG ATC CGA GAG TGC ATC CCT GCC CAC
Leu Pro Glu Phe Thr Ile Lys Glu Ile Arg Glu Cys Ile Pro Ala His
                                                360
                                                 *
TGC TTT GAG CGC TCC GGT CTC CGT GGT CTC TGC CAC GTT GCC ATC GAT
Cys Phe Glu Arg Ser Gly Leu Arg Gly Leu Cys His Val Ala Ile Asp
                                                            420
                                                             *
CTG ACT TGG GCG TCG CTC TTG TTC CTG GCT GCG ACC CAG ATC GAC AAG
Leu Thr Trp Ala Ser Leu Leu Phe Leu Ala Ala Thr Gln Ile Asp Lys

TTT GAG AAT CCC TTG ATC CGC TAT TTG GCC TGG CCT GTT TAC TGG ATC
Phe Glu Asn Pro Leu Ile Arg Tyr Leu Ala Trp Pro Val Tyr Trp Ile
```

FIG. 5A

```
                480
                 *
ATG CAG GGT ATT GTC TGC ACC GGT GTC TGG GTG CTG GCT CAC GAG TGT
Met Gln Gly Ile Val Cys Thr Gly Val Trp Val Leu Ala His Glu Cys

                                540
                                 *
GGT CAT CAG TCC TTC TCG ACC TCC AAG ACC CTC AAC AAC ACA GTT GGT
Gly His Gln Ser Phe Ser Thr Ser Lys Thr Leu Asn Asn Thr Val Gly

                                                600
                                                 *
TGG ATC TTG CAC TCG ATG CTC TTG GTC CCC TAC CAC TCC TGG AGA ATC
Trp Ile Leu His Ser Met Leu Leu Val Pro Tyr His Ser Trp Arg Ile

                                                            660
                                                              *
TCG CAC TCG AAG CAC CAC AAG GCC ACT GGC CAT ATG ACC AAG GAC CAG
Ser His Ser Lys His His Lys Ala Thr Gly His Met Thr Lys Asp Gln

GTC TTT GTG CCC AAG ACC CGC TCC CAG GTT GGC TTG CCT CCC AAG GAG
Val Phe Val Pro Lys Thr Arg Ser Gln Val Gly Leu Pro Pro Lys Glu

                720
                 *
AAC GCT GCT GCT GCC GTT CAG GAG GAG GAC ATG TCC GTG CAC CTG GAT
Asn Ala Ala Ala Ala Val Gln Glu Glu Asp Met Ser Val His Leu Asp

                                780
                                 *
GAG GAG GCT CCC ATT GTG ACT TTG TTC TGG ATG GTG ATC CAG TTC TTG
Glu Glu Ala Pro Ile Val Thr Leu Phe Trp Met Val Ile Gln Phe Leu

                                                840
                                                 *
TTC GGA TGG CCC GCG TAC CTG ATT ATG AAC GCC TCT GGC CAA GAC TAC
Phe Gly Trp Pro Ala Tyr Leu Ile Met Asn Ala Ser Gly Gln Asp Tyr
```

FIG. 5B

900

GGC CGC TGG ACC TCG CAC TTC CAC ACG TAC TCG CCC ATC TTT GAG CCC
Gly Arg Trp Thr Ser His Phe His Thr Tyr Ser Pro Ile Phe Glu Pro

CGC AAC TTT TTC GAC ATT ATT ATC TCG GAC CTC GGT GTG TTG GCT GCC
Arg Asn Phe Phe Asp Ile Ile Ile Ser Asp Leu Gly Val Leu Ala Ala

960

CTC GGT GCC CTG ATC TAT GCC TCC ATG CAG TTG TCG CTC TTG ACC GTC
Leu Gly Ala Leu Ile Tyr Ala Ser Met Gln Leu Ser Leu Leu Thr Val

1020

ACC AAG TAC TAT ATT GTC CCC TAC CTC TTT GTC AAC TTT TGG TTG GTC
Thr Lys Tyr Tyr Ile Val Pro Tyr Leu Phe Val Asn Phe Trp Leu Val

1080

CTG ATC ACC TTC TTG CAG CAC ACC GAT CCC AAG CTG CCC CAT TAC CGC
Leu Ile Thr Phe Leu Gln His Thr Asp Pro Lys Leu Pro His Tyr Arg

1140

GAG GGT GCC TGG AAT TTC CAG CGT GGA GCT CTT TGC ACC GTT GAC CGC
Glu Gly Ala Trp Asn Phe Gln Arg Gly Ala Leu Cys Thr Val Asp Arg

TCG TTT GGC AAG TTC TTG GAC CAT ATG TTC CAC GGC ATT GTC CAC ACC
Ser Phe Gly Lys Phe Leu Asp His Met Phe His Gly Ile Val His Thr

1200

CAT GTG GCC CAT CAC TTG TTC TCG CAA ATG CCG TTC TAC CAT GCT GAG
His Val Ala His His Leu Phe Ser Gln Met Pro Phe Tyr His Ala Glu

FIG. 5C

```
                              1260
                                *
GAA GCT ACC TAT CAT CTC AAG AAA CTG CTG GGA GAG TAC TAT GTG TAC
Glu Ala Thr Tyr His Leu Lys Lys Leu Leu Gly Glu Tyr Tyr Val Tyr

                                                1320
                                                  *
GAC CCA TCC CCG ATC GTC GTT GCG GTC TGG AGG TCG TTC CGT GAG TGC
Asp Pro Ser Pro Ile Val Val Ala Val Trp Arg Ser Phe Arg Glu Cys

                                                            1380
                                                               *
CGA TTC GTG GAG GAT CAG GGA GAC GTG GTC TTT TTC AAG AAG TAAAAA
Arg Phe Val Glu Asp Gln Gly Asp Val Val Phe Phe Lys Lys

                                                               1440
                                                                  *
AAAAGACAAT GGACCACACA CAACCTTGTC TCTACAGACC TACGTATCAT GTAGCCATAC

CACTTCATAA AAGAACATGA GCTCTAGAGG CGTGTCATTC GCGCCTCC
```

FIG. 5D

FastA Match of ma29 and contig 253538a

```
SCORES       Init1:    117 Initn:    225 Opt:    256
Smith-Waterman score: 408;     27.0% identity in 441 aa overlap

                        10        20        30        40        50
ma29gcg.pep   MGTDQGKT---FTWEELAAHNTKDDLLLAIRGRVYDVTKFLSRHPGGVDTLLLGAGRDVT
                  || |    |||:|:| ::  ::   |:|  :||::::|  |||||  ::   ||:|:|
253538a           QGPTPRYFTWDEVAQRSGCEERWLVIDRKVYNISEFTRRHPGGSRVISHYAGQDAT
                          10        20        30        40        50

               60        70        80        90       100       110
ma29gcg.pep   PVFEMYHAF-GAADAIMKKYYVGTLVSNELPIFPEPTVFHKTIKTRVEGYFTDRNIDPKN
                 |   :|    |  :     |::   :|  |  |  |  | |  |||  ::        |   | :  :   :
253538a       DPFVAFHINKGLVKKYMNSLLIGEL-SPEQPSF-EPTKNKELTDEFRELRATVERMGLMK
               60        70        80         90       100       110

               120       130       140       150        160       170
ma29gcg.pep   RPEIWGRYALIFGSLIASYYAQLFVPFVVERTWLQVVF-AIIMGFACAQVGLNPLHDASH
               :::   : | :   ::      |   :: :|    ::|   :: |:::: : ||:|     || :|
253538a       ANHVF--FLLYLLHILLLDGAAWLTLWVFGTSFLPFLLCAVLLSAVQAQAGWLQ-HDYGH
                  120       130       140       150       160       170

                180       190           200       210       220
ma29gcg.pep   FSVTHNPTVWKILGATHDF----FNGASYLVWMYQHMLGHHPYTNIAGADPDVSTSE---
              :|| ::|  |: |   :| |      ::|||   | ::| : ||    ||    ||||:   :
253538a       LSVYRKPK-WNHL--VHKFVIGHLKGASANWWNHRH-FQHHAKPNIFHKDPDVNMLHVFV
                   180        190        200        210       220

                 230       240        250         260       270       280
ma29gcg.pep   ----PDVRRIKPNQKWF-VNHINQHMFV--PFLYGLLAFKVRIQDINILYFVKTNDAIRV
                    ::  | : |::  || ::::|:  | |     : |: :|     |: ::  :: : :
253538a       LGEWQPIEYGKKKLKYLPYNHQHEYFFLIGPPLLIPMYFQYQI----IMTMIVHKNWVDL
              230        240       250       260       270           280

                  290       300       310       320       330       340
ma29gcg.pep   NPISTWHTVMFWGGKAFFVWYRLIVPLQYLPLGKVLLLFTVADMVSSYWLALTFQANHVV
                  :| : ::      ||: |    :|: |   || :||:::    :: |:|:: : | ||:|
253538a       ----AWAVSYYI---RFFITY---IPF-YGILG-ALLFLNFIRFLESHWFVWVTQMNHIV
                   290             300           310       320       330

                  350       360            370       380       390
ma29gcg.pep   EEVQWPLPDENGIIQKDWAAMQVETT----QDYAHDSHLWTSITGSLNYQAVHHLFPNVS
              |:        |:::    :||  : |: :|      |:: :|    |  |     | ||:|   |||||::
253538a       MEI-----DQEAY--RDWFSSQLTATCNVEQSFFND---WFS--GHLNFQIEHHLFPTMP
                          340       350       360                 370

              400       410       420       430       440
ma29gcg.pep   QHHYPDILAIIKNTCSEYKVPYLVKDTFWQAFASHLEHLRVLGLRPKEEX
              :|:     |   ::|:  |::: : |   |
253538a       RHNLHKIAPLVKSLCAKHGIEYQEKPLLRALLDIIRSLKKSGKLWLDAYLHKX
            380       390       400       410       420       430
```

Figure 9